

# English Proficiency Test

## THE ORAL COMPONENT OF A PRIMARY SCHOOL

MANY TEACHERS FEEL COMFORTABLE SETTING PENCIL-AND-PAPER TESTS. YEARS OF experience marking written work have made them familiar with the level of written competence pupils need in order to succeed in a specific standard. However, teachers often feel much less secure when dealing with tests which measure speaking and listening even though these skills are regarded as essential components of a diagnostic test which measures overall linguistic proficiency. Although the second-language English pupils often come from an oral rather than a written culture, and so are likely to be more proficient in this mode of communication, at least in their own language, speaking in English may be a different matter. In English medium schools in particular a low level of English may impede students' acquisition of knowledge. Therefore, identifying the correct level of English of the student is all the more challenging and important.

This article outlines some of the problem areas described by researchers when designing a test of oral production for beginning-level speakers of English and suggests ways in which they may be addressed.

**How does one set a test which does not intimidate children but encourages them to provide an accurate picture of their oral ability?**

In replying to this question, one needs to consider briefly the findings of researchers working in the field of language testing. "The testing of speaking is widely regarded as the most challenging of all language tests to prepare, administer and score," writes Harold Madsen, an international expert on testing (Madsen 1983:147). This is especially true when examining beginning-level pupils who have just started to acquire English, such as those applying for admission to primary school. Theorists suggest three reasons why this type of test is so different from more conventional types of tests.

Firstly, the nature of the speaking skill itself is difficult to define. Because of this, it is not easy to establish criteria to evaluate a speaking test. Is "fluency" more important than "accuracy," for example? If we agree fluency is more important, then how will we define this concept? Are we going to use "amount of information conveyed per minute" or "quickness of response" as our definition of fluency?

A second set of problems emerges when testing beginning-level speakers of English, which involves getting them to speak in the first place, and then defining the role the tester will play while the speaking is taking place. Relevant elicitation procedures which will prompt speakers to demonstrate their optimum oral performance are unique to each group of speakers and perhaps even unique to each occasion in which they are tested. The tester will therefore need to act as a partner in the production process, while at the same time evaluating a number of things about this production.

A third set of difficulties emerges if one tries to treat an oral test like any other more conventional test. "In the latter, the test is often seen as an object with an identity and purpose of its own, and the children taking the test are often reduced to subjects whose only role is to react to the test instrument" (Madsen 1983:159). In oral tests, however, the priority is reversed. The people involved are important, not the test, and what goes on between tester and testee may have an existence independent of the test instrument and still remain a valid response.

**How can one accommodate these difficulties and still come up with a valid test of oral production?**

In answering this question, especially in relation to the primary school mentioned earlier, I would like to refer to the experience I and one of my colleagues, Viv Linington, had in designing such a test for the Open Learning Systems Education Trust (OLSET) to measure the success of their English-in-Action Programme with Sub B pupils. This Programme is designed to teach English to pupils in the earliest grades of primary school, using the medium of the tape recorder or radio.

In devising this test, we decided to use fluency as our basic criterion, i.e., "fluency" in the sense Brumfit uses it: "the maximally effective operation of the language system so far acquired by the student" (Brumfit 1984: 543). To this end, we decided to record the total number of words used by each pupil on the test administration and to employ this as an overall index to rank order the testees in terms of performance.

To address the second and third set of problems outlined above, we decided to use elicitation procedures with which the children were familiar. Figures 1 and 2 would require the teacher to find a picture full of images the pupils could relate to such as children playing. Students could participate in the following types of activities:

- an informal interview, to put the children at ease by getting them to talk about themselves, their families and their home or school lives (See Figure 1).
- a set of guided answers to questions about a poster, to test their knowledge of the real life objects and activities depicted on the poster as well as their ability to predict the consequences of these activities (See Figure 2).
- narratives based upon packs of story cards, to generate extended language in which the children might display such features as cohesion or a knowledge of the English tense system in an uninterrupted flow of speaking.

Instead of treating the situation as a "test," we asked testers to treat it as a "game." Both partners would be seated informally on the ground (with, in our case, a recorder placed

## THE INTERVIEW

The tester should capture personal details by asking the following type of questions:

What is your name?  
 Where do you live?  
 Do you have any brothers or sisters?  
 Does anyone else live at home with you?  
 Now tell me, what do you all do when you get up in the morning?  
 How do you all go to school and work?  
 Do you have any brothers or sisters in this school?  
 What standards are they in?  
 Which subject do you enjoy most? Why?  
 What do you do at break?  
 Tell me about your best friends.  
 What does your mother/grandmother cook for dinner?  
 Can you tell me how she cooks it?  
 Why do you all enjoy this food most?  
 Do you listen to the radio/watch TV in your house?  
 What is your favorite programme?  
 Why do you enjoy it most?  
 What do you do when you are getting ready to sleep in the evening?  
 What time do you go to sleep. Why?  
 Now look at the picture and tell me what this little boy is doing. Let's give him a name.  
 What do you suggest?

FIGURE 1

unobtrusively on the floor between them because of the research nature of our test). If the occasion was unthreatening to the pupil with the tester acting in a warm friendly way, we anticipated the child would respond in a similar way, and thus produce a more accurate picture of his or her oral productive ability. We suggested the tester act as a Listener/Speaker only while the test was being conducted, and as Assessor once the test administration was over.

To maintain a more human approach to the testing situation, we decided to allow the tester a certain flexibility in choosing questions to suit each particular child, and also in the amount of time she spent on each subtest. The time allowed for testing each pupil would be limited to 8 minutes, and all three subtests would be covered during this period, but the amount of time spent on each could vary.

Question banks were provided for testers to select questions they felt were within the range

of each child's experience, but there was an understanding that *how* and *why* questions were more difficult to answer than other *Wh*-questions. A range of both types should therefore be used.

Story packs also provided for a range of experiences and could be used by the tester telling a story herself first, thus demonstrating what was required of the pupil. However, it was anticipated that some pupils might be sufficiently competent to use the story packs without any prompting from the teacher. Pupils could place the cards in any order they chose, as the sole purpose of this procedure was to generate language. Story packs were composed of picture stories that had been photocopied from appropriate level books, cut up into individual pictures, and mounted on cardboard. Six pictures to a story pack were considered sufficient to prompt the anticipated length of a story pupils could handle.

This test of oral production was administered at both rural and urban schools to children who were on the English-in-Action Programme and those who were not. The comparative results are not relevant here, but findings about which aspects of the test worked and which did not may be of assistance to those who wish to set similar tests. In summarising these findings, I will comment on the administration of the test, the success of each subtest in eliciting

## "ARRIVING AT SCHOOL"

Questions for guided response:

What are the children doing?  
 Where are they?  
 How many children are there?  
 Are there more boys than girls?  
 How do you know this?  
 What is the girl in the green dress doing?  
 What are the boys going to do when they finish playing marbles?  
 Do you think the children are happy?  
 Have you ever played marbles?  
 (If yes) How do you play marbles?  
 (If no) What other game do you play with your friends?  
 How do you play it?

Now look at the picture and tell me what this little boy is doing. Let's give him a name.  
 What do you suggest?

FIGURE 2

language, and, finally, on the criteria we used for evaluating the test outcomes.

Firstly, both testers commented that this type of test was more difficult to organise and administer than other kinds of evaluation tests they had used. This was caused by the need to find a quiet and relatively private place to administer the test and record the outcome and because the procedure could be done only on a one-to-one basis. We had anticipated this type of feedback but were also not surprised when told that subsequent administrations “were much easier and the children were more enthusiastic about participating than the previous time.” The testing procedure was new to both tester and testee, but once experienced, it gave children greater freedom of expression than other kinds of tests.

Secondly, while the test as a whole did elicit oral language production, the amount and type of language varied from subtest to subtest. The interview produced rather less language than the other two subtests; it also elicited rather learned chunks of language, which we called “patterned responses.”

The guided responses, on the other hand, produced a much greater variety of answers, couched in a fairly wide range of grammatical structures. But even these responses consisted on the whole of single words or phrases. Open-ended questions evoked longer responses from the more able students, but seemed to confound less able students. For example, the question “What can you see in the picture?” produced the answer “I can see a car and a woman going to the shop and a boy had a bicycle and the other one riding a bicycle,” from a bright pupil, but only “Boy and bicycle” from a weaker pupil.

Higher order *Wh*- questions such as “What do you think is in the suitcase?” or “What will happen next?” seemed to produce only “I don’t know” responses from even the most competent pupils. They seemed to lack the linguistic resources, or perhaps the cognitive resources, to predict or suggest answers.

The narrative subtest, based on the story cards, elicited the best display of linguistic ability from the testees, both in terms of amount of language produced and range of grammatical structures used.

Competent pupils were able to respond well to the tell/retell aspect and constructed sentences of 7 to 10 words in length, joined by

a variety of coordinating devices. They also employed past tense forms in retelling the story such as the following:

The boys they played with the cow’s what  
..... what ..... a ..... bells three bells  
..... then they got some apples and went  
to swim ..... the monkey saw them swim  
and putted them shirts and shorts .....  
some they said hey ..... I want my shirts  
..... wait I want my shirts ..... but mon-  
key she run away

Less competent students could describe isolated images on each card without using narrative in any way to link them together.

From these results we therefore concluded that the story packs were the most successful of the three elicitation procedures we used in stimulating optimum language output.

The final issue from the findings of the OLSET test that are relevant here are the criteria used for assessing the language output. Our decision to count “number of words produced” as a measure of speaking ability was a mixed blessing. Initially it did seem to rank order the pupils in terms of ability and gave us a base for comparison at subsequent test administrations, but non-verbal factors such as self confidence, familiarity with the tester, and presence of the teacher may have affected even these results. In the second administration of the test, it was not at all accurate because improvement in ability to speak and respond in English was reflected more in the quality of how the testees spoke, rather than in the quantity of language they produced. Several of the more competent pupils spoke the second time in round 1 but displayed knowledge/features not present in their own home languages such as prepositions and articles, used correctly subordinating and coordinating conjunctions they had been introduced to only in the course of conversation, and employed a variety of tenses in their story telling. We therefore used this data to develop a number of assessment levels, or descriptive band scales, based upon these various grammatical competencies, when evaluating the pupils’ output (a band scale outlines a set of linguistic features and skills a pupil needs to display in order to be placed in that category).

In response to our discussion, some schools have begun to introduce two components in their diagnostic test. The first is a multiple-

## English Proficiency Test... | Hingle and Linington

*continued from page 33*

choice comprehension test and the second an oral test based upon a set of story cards.

The same test will be used for pupils at all levels of the primary school, using the lead provided by a test produced by the Human Sciences Research Council for the same purpose. However, the expected proficiency levels to enter a particular grade or standard will be different.

In conclusion, let me summarise the advice I would give to teachers who need to design speaking tests but who are afraid to take the plunge into this area of assessment:

- Do not be afraid to set such a test in the first place.
- Draw on your own materials to set a test appropriate for your group of testees.
- Keep the factor of time constant for each test administration.

- Give the testee the opportunity to lead once he or she is at ease.
- Do not allow factors such as accent to cloud your perception of linguistic competence.
- Rely on your own instinctive judgment when assigning a value to performance on such a test.
- Try and think of this value in terms of words rather than marks.

### References

- Brumfit, C. 1984. *Communicative methodology in language teaching*. Cambridge: Cambridge University Press.
- Madsen, H. S. 1983. *Techniques in testing*. New York: Oxford University Press.

This article was originally published in the April 1997 issue.

